Computer Engineering and Applications Vol. 5, No. 3, October 2016

Boosting Diagnosis Accuracy of Alzheimer's Disease Using Statistical and Kernel-Based Feature Selection Techniques

M. A. Balafar[1]*, Rouya Norzadeh[2]

1Dept. of IT, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran; balafarila@tabrizu.ac.ir

2Dep. Of Computer, Azad univ of Ghermi, Ghermi, Iran; norzadeh.rouya@gmail.com

ABSTRACT

Alzheimer's disease (AD) is a popular disease in the elderly. Approximately, 26 million people worldwide are affected by AD. Among the various diagnostic methods for Alzheimer's disease, MRI brain imaging can display sharp changes in brain tissues. It can be used as a medium for early Alzheimer detection. Considering the high volume of features related to brain tissue thickness, requires the using feature reduction methods. For this purpose, statistical tests pair sample test and Independent sample test was used. After careful selection of key features, for reducing the number of features, SAS which is a kernel-based feature selection algorithm is used in linear and nonlinear mode. At the end, neural network classification, decision trees, nearest neighbor and Naïve Bayes algorithms are used for modeling. Results show that the classification accuracy of obtained feature subsets have better results compare to the original data set.

Keywords*:* Feature Selection, Classification Accuracy, Neural Networks, Decision Trees

1. INTRODUCTION

Alzheimer detection is getting progressively more important. Different methods are proposed to improve accuracy of Alzheimer detection. There are great success in this area, but, still there is challenging to solve. Researchers are working to overcome Alzheimer detection challenges. There are different areas in this subject to be work on. Some of researchers try to improve learning models and others try to use new features and new combination of features [1]. Unimodal data yields week results, therefore, data fusion of several modal data are used to boost accuracy of Alzheimer detection [2]. Also, fusing of different class of features is used to improve Alzheimer detection accuracy [3]. These features could be from different modalities or from different class of features [4]. In this paper, a two phase feature selection algorithm is utilized to boost accuracy of Alzheimer detection. In the first phase, statistical tests are used to select features, then, kernel-based feature selection algorithm is utilized. In the second phase, the algorithms are used for first time in Alzheimer detection.

Various sources in Alzheimer detection using brain tissue thickness characteristics have been studied. Most studies on Alzheimer detection are focused on four groups of people: healthy subjects (NC), Alzheimer's disease (AD), those with mild disease and later recovered (MCI -S) and those with mild disease later become Alzheimer's disease (MCI-P) [5]. Most of the studies used the statistical

methods for diagnosis. In this paper, the SAS, a kernel-based feature selection algorithm, is used, for the first time in the field of Alzheimer's disease. The goal of this article is to select characteristics features of brain tissue thickness extracted from MRI images for classification and separation of Alzheimer's disease and healthy individuals. For this purpose, the study dataset was obtained from ADNI website. After normalizing the data, a subset of characteristic feature is selected in two steps. First using statistical test, then, using kernel-based feature selection algorithm. For this purpose in according to the behavior of the algorithm, 10 subset of features respectively with 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 features are selected. Then, the accuracy of classification of the subsets was evaluated using neural network and decision trees classification algorithms. The purpose of this article, in addition to select a subset of features, is specifying the effect of the number of features on the accuracy of classification.

2. PREPARATION DATASET

The study dataset has been downloaded from ADNI website (http://adni.loni.usc.edu/data-samples/access-data). The goal of this database is to collect research data and made them available to researchers. Dataset consists of 50 samples, 27 AD patient people and 23 healthy people and there is 16300 properties per sample. In this dataset, people during periods of the baseline, 6 months, 12 months, 24 months and 36 months were studied. Finally, there is a dataset with 50 samples and 440 features per period.

3. DATA NORMALIZATION

There are a lot of methods for data normalization. In all of standardization methods, data is converted in a manner that the conditions to be met. In order to implement the classification algorithms such as neural networks, the data values should be normalized. For this purpose, all data is scaled to the range of [200 and 100]. For this purpose, The MIN-MAX normalization method is used. In this normalization method, assuming that the data Z is in the [Z_{min} ,Z_{max}] interval, the goal is to change range to the new interval [New_{min} ,New_{max}]. For this purpose, each initial data value such as V will convert to the new value $\acute{V}$ in the new interval. The normalization equation is given in Equation (1). [6].

$$Z_i = \frac{V-(Z_{min})}{(Z_{max})-(Z_{min})} = \frac{\acute{V} - New_{min}}{(New_{max} - New_{min})} \quad (1)$$

4. REDUCING NUMBER OF FEATURES USING STATISTICAL TESTS

Using statistical methods, the optimum features will be selected. 880 features are available in the dataset studied for two periods. In other word, there are 440 features per period. The features with low effect in discrimination of AD and healthy subjects could be eliminated. Due to the high volume of features, it is better to use the statistical methods that are fast in determining the low importance features. For this purpose, the t test in SPSS software is used.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

4.1 TEST PAIRED SAMPLE TEST

With this test, the following question could be answered. Is there any difference between any of the features of the brain tissue thickness at the baseline and 24 months periods or not?

If there is such a difference, these features are important and such important features should be selected. If there is no significant difference between values of any feature in the two periods, that feature has low importance in classifying two groups of patients and healthy individuals. This feature can be eliminated.

The data of these two periods are arranged as ordered pairs. The test statistics are based on the variable di which is difference of each pair. If the measured variable in the first period be xi and the measured variable in the second period be yi,the value of di will be obtained as follows (Adel. et al., 1379) [7].

$$d_i = x_i - y_i \quad (2)$$

Then, these variables, assuming normal distribution for data and having unknown variance, follows the distribution of paired t test, as Eq (3).

$$t = \frac{d}{sd/\sqrt{n}} \quad (3)$$

Where the sd value is obtained from Eq (4).

$$sd = \sqrt{\frac{1}{n-1}\sum(d_i - \bar{d})} \quad (4)$$

In Eq(4), d is standard deviation and $\bar{d}$ is the average of d. Here, the null hypothesis is having no difference between couples which is zero and 1 hypothesis shows a difference. Test results will confirm or reject the null hypothesis.

To apply this test it is necessary to determine the two main variables, Confidence interval and P-value. Given that used dataset is a medical dataset and has high sensitivity, the value of confidence interval is fixed to 99% instead of 95%. This improves the accuracy of the feature selection (Davies et al., 2009) [8].

4.2 INDEPENDENT SAMPLE TEST

This test is used to get the average of two independent groups that are organized separately. The equation (5) is used for this purpose, where the difference between the mean of the two groups has been considered according to the variance and size of the target groups. In this equation, $\overline{x_1}$ and $\overline{x_2}$ are the mean of groups one and two. n1 and n2 are the number of members of groups one and two. s1 and s2 are the variance in groups one and two which are obtained using equation (6). Independent statistical hypothesis for the independent sample test comes as (7). In this case, if μ1 and μ2 are the average of two groups of healthy and patient, the null hypothesis is that there is no significant difference between the two groups and one hypothesis is that there is a significant difference between the two groups.

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left[\frac{n_1+n_2}{n_1 n_2}\right]}} \quad (5)$$

$$s_i = \frac{\sum(x_i - \bar{x}_i)}{n_i - 1} \quad (6)$$

$$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2 \qquad (7)$$

5. KERNEL-BASED FEATURE SELECTION ALGORITHMS

After reducing the number of features by statistical test, a subset of important features is selected using linear and non-linear feature selection algorithms. For linear algorithms for linear kernel are used. When the data placed in linear space, this kernel has better performance. For the implementation of nonlinear algorithms different kernels can be used that include: dot, quadratic, polynomial and Gaussian [9]. Generally radial Gaussian kernel function (RBF) has better performance in nonlinear prediction problems. This kernel is one of the most-used kernels in the field of machine learning algorithms. This kernel is used in the most kernel-based machine learning algorithms. By increasing the dimension of features, RBF reveals the hidden relationships between the features. This kernel is obtained using the equation (8).

$$K(x, y) = \exp\left(-\frac{\|x_i - y_i\|}{2\sigma^2}\right) \qquad (8)$$

5.1 SCALED ALIGNMENT SELECTION ALGORITHM

The SAS algorithm is one of kernel-based feature selection algorithms (Ramona et all., 2012). It reduces the number of features by weighting the features by a decision function. This is the used algorithm to select a subset of features [10]. This approach was based on the KPO search strategy. The algorithm is shown in Figure 1:

Algorithm 1 Method 1: SAS

$\theta_0 = [1, \ldots, 1], \quad \mathcal{A}_0 = 0, \quad n = 0$
$\boldsymbol{K}^* = \boldsymbol{y}\boldsymbol{y}^T$
repeat
  $n \leftarrow n + 1$
  Gram-matrix computation: $[\boldsymbol{K}_{\boldsymbol{\theta}}]_{ij} = k_\theta(\boldsymbol{x}_i, \boldsymbol{x}_j)$
  Parameters update: $\boldsymbol{\theta}_n \leftarrow \boldsymbol{\theta}_{n-1} + \eta\, \hat{c}_{\boldsymbol{\theta}} \mathcal{A}_{n-1}(\boldsymbol{K}_{\boldsymbol{\theta}}, \boldsymbol{K}^*)$
until convergence: $|\mathcal{A}_n - \mathcal{A}_{n-1}| < \epsilon$

FIGURE 1: Algorithm SAS [10].

Here, using this algorithm, subsets with feature numbers: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 are obtained.

6. THE CLASSIFICATION ALGORITHMS

Classification is to assign records or any collection of objects to a specified set of classes (categories). Classification has many applications in natural language processing tasks such as e-mail filtering and identifying user's tendencies, etc.

In this paper, the classification accuracy of obtained subsets is compared using following classification algorithms.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

6.1 NEURAL NETWORK CLASSIFICATION ALGORITHM

Neural networks are systems and computational methods for machine learning, knowledge representation and the prediction the output of complex systems using gained knowledge. The main idea of these networks is inspired from biological nervous system function. They process data and information in order to learn existing pattern and create related knowledge. In other words, they create new structures to process the information [11]. Neural network architecture is formed of input layer, hidden layer and output layer. The number of these layers will be determined by the designer through trial and error in order to reduce the error rate. In this paper, 90% of data is for testing and 10% of data for teaching and learning.

6.2 DECISION TREE CLASSIFIER

Decision tree is one of the common methods which are used for inductive inference. In fact, it is useful for issues that there is a single answer in form of a class or category. For example, the decision tree can answer the question: Is a person an Alzheimer's disease patient or not. This method based on the training data at any stage, select a feature. Then, using that feature, the data set is divided into two or more categories. This continues until all the data in a category belong to a class. The goal is to build a tree of features that have the lowest height in order to searches on fewer features for label test data [11].

6.3 K-NEAREST NEIGHBOR CLASSIFICATION ALGORITHM

One of the best classifiers is K nearest neighbor classifier. KNN classification due to having no need to make assumptions on the data is a simple and widely used classifier. This classifier considers that the test sample belongs to the class that has the most votes in the k nearest neighbors. (Edenia et al., 2014) classification accuracy of this method depends on the amount of K. Due to small number of samples in studied dataset which is 50 the value of k is set to 10 and 7. Using these two determined k values, classification accuracy for selected subsets of features is evaluated, separately. Then, it is determined that with which K algorithm has more accurate results.

6.4 Naïve Bayes Classifier Algorithm

Naïve bayes algorithm is on one of the most efficient algorithms for some learning issues (Vafa et al., 2009) [12]. A complete comparison is done between this algorithm and other algorithms, such as decision trees and neural network learning (Mychy, 1994) [13-14]. The researcher found that the Naïve bayes classification algorithm in is better than the other algorithms in some cases.

7. RESULTS

To perform statistical tests including independent t test and two-sample t-test, SPSS statistical software version 18 is used. First, features that are important for the classification of Alzheimer's disease and healthy people are selected and stored in an Excel file. The data normalization code is written and run in the MATLAB version 2013. The SAS feature selection algorithm in the MATLAB software version 2013 is used. The output of this algorithm (selected features) saved in Excel file format. To evaluate the classification accuracy of classification algorithms (neural networks, decision trees, Naïve Bayes on, k-nearest neighbour) with selected features, Rapid Miner software has been used.

In this section results are represent in the graphs and tables format. In Figure 2 the results of the implementation of the Pair Sample Test algorithm is shown. As can be seen, the number of features for each period has been reduced from 440 to 141 features. In Figure 3, the results of the Independent Sample Test algorithm are shown. As can be seen, the number of features is reduced from 141 to 106. Until now, using these two tests, the number of minor properties has been significantly reduced.

Using the filtering method, 106 features were chosen from among 440 attributes for each period. At this stage, dataset contains 50 samples (n = 27 AD patients and 23 healthy subjects) and 106 features. This dataset is used as input to the SAS algorithm. Using kernel-based feature selection algorithms with two separate kernels, from a total of 106 features, 10 separate subset of the features of brain tissue thickness with a number of features 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 are obtained.

The purpose of specifying subsets with different number of features is to investigate the effect of selected features and number of them on the classification accuracy.

At last, the classification accuracy using the selected subsets of features with different classification algorithms: Naïve Bayes, neural networks, nearest neighbour and decision tree algorithms, are compared. As can be seen in Figures 4 and 5, the best accuracy is obtained for subsets with 20, 30, 40 and 50 features. The Naïve Bayes algorithm classifier has the best performance.

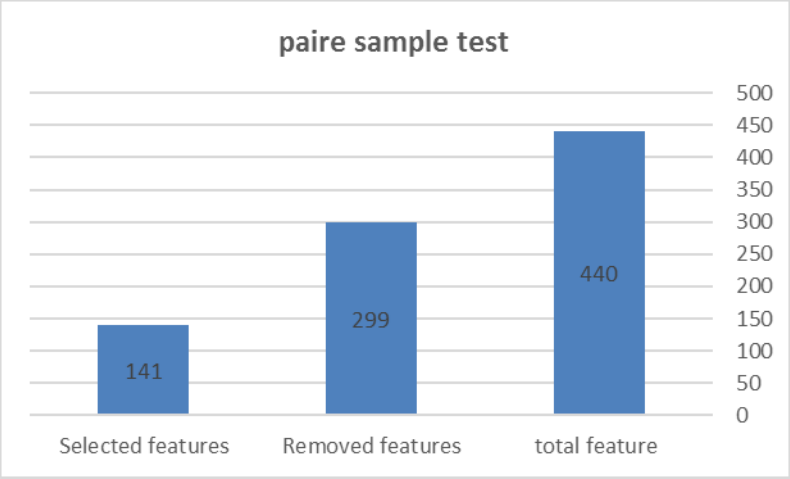

FIGURE 2. The results of algorithm PAIR SAMPLE TEST

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

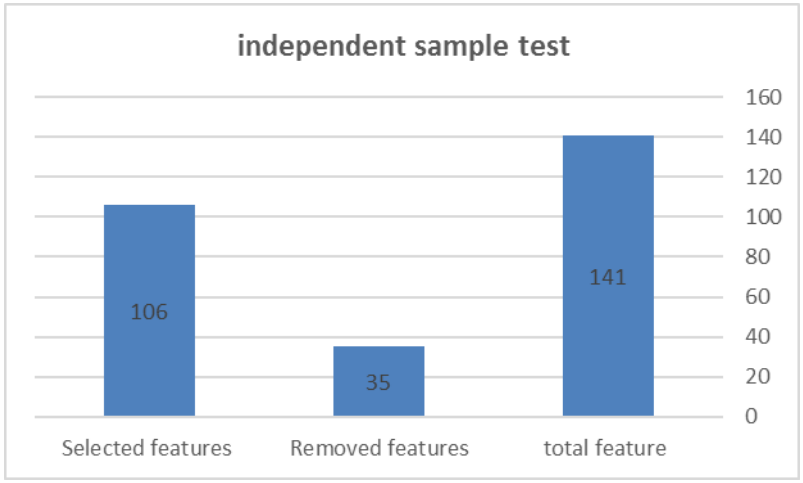

FIGURE 3. The results of algorithm INDEPENDENT SAMPLE TEST

Figure 4 presents the best accuracy, when sasliner is used for feature selection, which are as follows: by decision tree for a subset of 10 features, accuracy is 90%; for subset of 30 features and knn algorithm with k = 7, accuracy is 98%; for subset of 30 features and knn algorithm with k = 10, accuracy is 96%; for subsets of 30 and 20 features and Naïve Bayes algorithm, accuracy is 98%; and, for subsets of 40 and 50 and neural network algorithm, accuracy is 98%. As a result, the subsets of 20 or 30 features with Naïve Bayes algorithm give the reasonable results. Figure 5 presents the accuracy of feature selection using sasrbf algorithm. The best accuracy are as follows: accuracy with decision tree algorithm and 10 features is 84%; for subset of 20 features and knn algorithm with k = 7, accuracy is 92%; for a subset of 20 features and knn algorithm with k = 10, accuracy is 92%; for subsets of 10 and 20 features and Naïve Bayes algorithm, accuracy is 92%; and, for the subset of 20 features and neural network algorithm, accuracy is 90%.

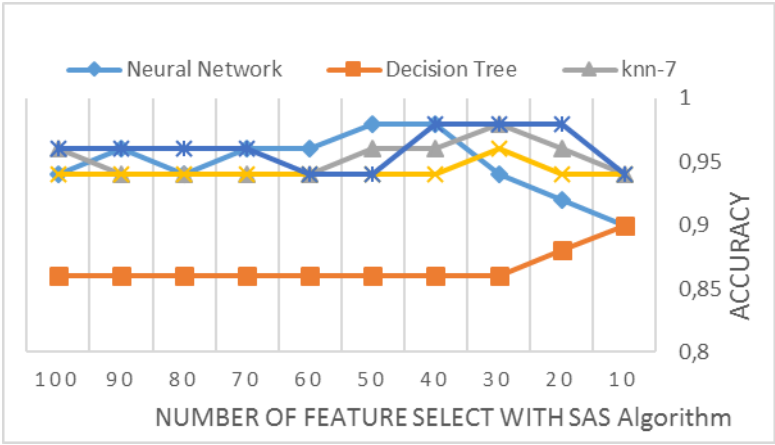

FIGURE 4. Comparison Chart of features selected by the accurate SASLINER algorithm.

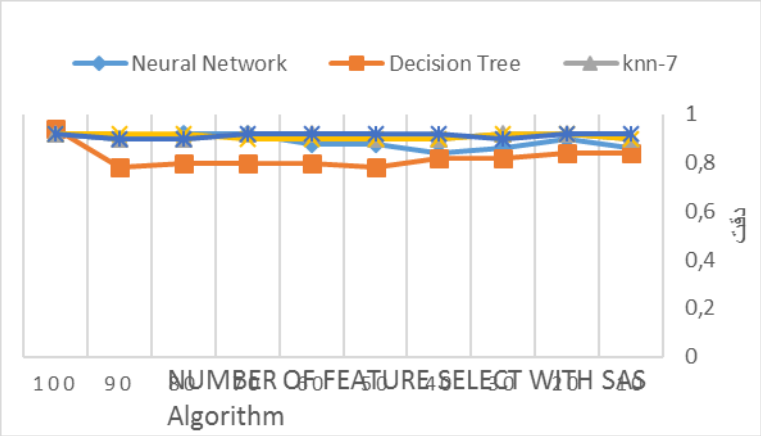

FIGURE 5. Comparing the number of features selected by the accurate SASRBF algorithm.

The proposed algorithm overcomes the work presented in the paper [15]. In the paper [15], Results indicate superiority of supervised machine learning techniques over unsupervised ones in diagnosing AD and withal, predominance of RBF kernel over lineal one. Accuracies of 83.3%, 83.3%, 90% and 91.7% are achieved in classification by K-mean, FCM, linear SVM and SVM with radial based function (RBF) respectively. In our postulated algorithm, best accuracy of 98%, 92% are achieved in classification by naïve bayes with lineal and radial based function (RBF), respectively.

8. CONCLUSION

MRI brain imaging is usually used for early diagnosis of Alzheimer's disease. MRI brain features of previous cases are used to learn a model for early diagnosis of AD. The main problem in this process is high volume of features. Therefore, in this paper, two feature selection algorithms are utilized which till now they have not been used in this area. For this purpose, first, statistical tests pair sample test and Independent sample test were used to eliminate low important features. Then, SAS is used in linear and nonlinear mode to reduce feature numbers more. To evaluate effeciency of feature selection, several classification have been used to classify the data with selected featuers. Results show that classification accuracy even improve after feature reduction.

REFERENCES

[1] Lei, B., Chen, S., Ni, D., & Wang, T. (2015, October). Joint learning of multiple longitudinal prediction models by exploring internal relations. In International Workshop on Machine Learning in Medical Imaging (pp. 330-337). Springer International Publishing.

[2] Liu M., Zhang D., Shen D. (2014). Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. Hum. Brain Mapp.35, 1305–1319.

[3] Yuan Y. H., Sun Q. S., Ge H. W. (2014). Fractional-order embedding canonical correlation analysis and its applications to multi-view dimensionality reduction and recognition. Pattern Recognit.47, 1411–1424

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[4] Yeh Y., Huang C., Wang Y. (2014). Heterogeneous domain adaptation and classification by exploiting the correlation subspace. IEEE Trans. Image Process.23, 2009–2018

[5] Mark D, Meadowcroft BS, et al., "MRI and histological analysis of beta-amyloid plaques in both human Alzheimer's disease and APP/PS1 transgenic mice", Journal of Magnetic Resonance Imaging, 29(5):997-1007, 2009.

[6] Bhandare S, Kumar Jain Y., "Min Max Normalization Based Data Perturbation Method for Privacy Protection", International Journal of Computer & communication Technology, 2(8):45-50, 2011.

[7] Adel A, Momeni M., Statistic and its application in management, Samt, pp. 283

[8] Davies H, Crombie L., "What are confidence intervals and p-values?", Hayward Medical Communications. Report NO: NPR09/1106.

[9] Ramona M, et al., "Multiclass feature selection with kernel Gram-matrix-based criteria", IEEE Trans Neural Netw Learn Syst, 23(10):1611-1623, 2012.

[10] Moghaddamnia A, et al., "Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques", Advances in Water Resources, 32:88–97, 2009.

[11] Quinlan J. R., "Induction of Decision Trees", Machine Learning, Springer, 1(1)81-106, 1986.

[12] Wafa S, et al., "Development of Genetic-based Machine Learning for Network Intrusion Detection", World Academy of Science, Engineering and Technology 55: 20-24, 2009.

[13] Magnin B, qe al., "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI", Neuroradiology, 51(2):73-83, 2009.

[14] Altheneyan A.S, et al., "Naive Bayes classifiers for authorship attribution of Arabic texts", Journal of King Saud University – Computer and Information Sciences, 26(4):473–484, 2014.

[15] Farzan A, et al., "Boosting diagnosis accuracy of Alzheimer's disease using high dimensional recognition of longitudinal brain atrophy patterns", Journal of Elsevier, 290: 124-130, 2015